



Human Re-Identification with a Robot Thermal Camera Using Entropy-Based Sampling

Serhan Coşar¹ · Nicola Bellotto¹

Received: 4 September 2018 / Accepted: 16 April 2019
© The Author(s) 2019

Abstract

Human re-identification is an important feature of domestic service robots, in particular for elderly monitoring and assistance, because it allows them to perform personalized tasks and human-robot interactions. However vision-based re-identification systems are subject to limitations due to human pose and poor lighting conditions. This paper presents a new re-identification method for service robots using thermal images. In robotic applications, as the number and size of thermal datasets is limited, it is hard to use approaches that require huge amount of training samples. We propose a re-identification system that can work using only a small amount of data. During training, we perform entropy-based sampling to obtain a thermal dictionary for each person. Then, a symbolic representation is produced by converting each video into sequences of dictionary elements. Finally, we train a classifier using this symbolic representation and geometric distribution within the new representation domain. The experiments are performed on a new thermal dataset for human re-identification, which includes various situations of human motion, poses and occlusion, and which is made publicly available for research purposes. The proposed approach has been tested on this dataset and its improvements over standard approaches have been demonstrated.

Keywords Service robots · Re-identification · Elderly care · Thermal camera · Occlusion · Body motion

1 Introduction

The ageing population and increased life-expectancy of people worldwide motivated the growing number of well-being and health monitoring applications for personal and domestic use. Service robotics is a promising research field that contributes to the creation of new solutions for elderly care. In recent years, service robots have become very popular by accomplishing various tasks, from guiding visitors in public environments to assisting elderly people at home. For the latter, in particular, a robust human re-identification system is needed in order for the robot to

distinguish between two or more users in the household and provide personalised services (e.g. medication reminders).

Considering its importance, human re-identification has not been sufficiently investigated in robotic applications. A very large amount of work focused on recognizing people across a network of RGB cameras in surveillance systems [2, 26]. In most of these applications, re-identification is performed by extracting appearance features from RGB images [5, 10, 16]. On the other hand, by exploiting RGB-D cameras, anthropometric features (e.g., limb lengths) extracted from skeleton data [1, 21], point cloud information [20] and volumetric features extracted from depth image [8] can be used for re-identification in service robot applications.

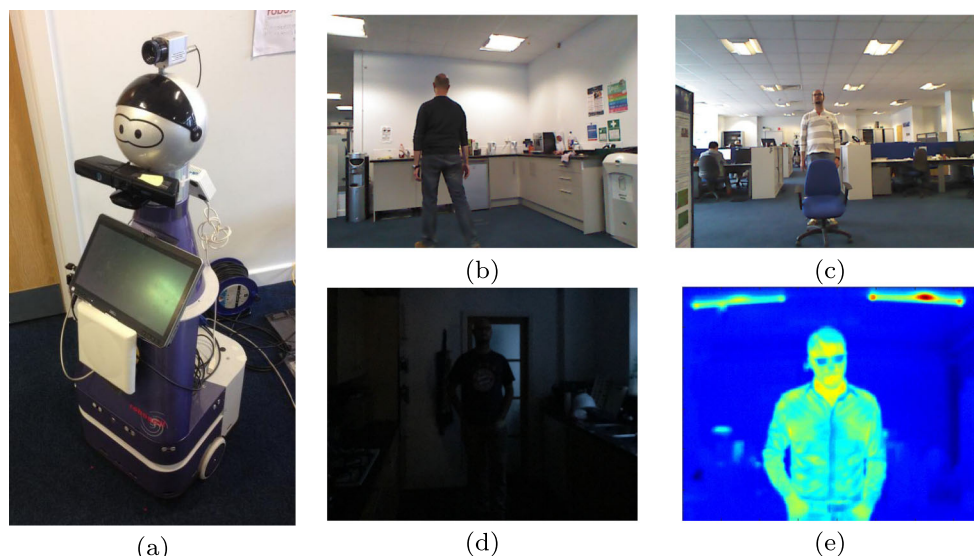
However, for long-term applications of domestic service robots, many existing approaches have strong limitations. For instance, appearance-based approaches are not applicable as people change often their clothes. In addition, for poorly illuminated or dark environments (at night), which are typical in domestic environments, RGB images provide very little information (Fig. 1d). Skeletal data is not always available because of self-occluding body motion (e.g. person facing opposite the camera, Fig. 1b) or objects occluding parts of the body (e.g., passing behind a table, see Fig. 1c).

✉ Serhan Coşar
scosar@lincoln.ac.uk

Nicola Bellotto
nbellotto@lincoln.ac.uk

¹ Lincoln Centre for Autonomous Systems (L-CAS),
School of Computer Science, University of Lincoln,
LN6 7TS Lincoln, UK

Fig. 1 Examples of human observed by a Kompi service robot **a** from the back **b**, under occlusion **c**, with poor lighting conditions **d** and on thermal camera **e**



In order to deal with the above limitations, in this paper we propose the use of a thermal camera, which provides clear images in the infrared spectrum, even in the darkness (Fig. 1e). The camera is mounted on the top of an interactive service robot (Fig. 1a) used in the ENRICHME¹ project to monitor and assist elderly people with mild cognitive impairments at home.

In this kind of robotics applications, it is very hard to collect large amount of thermal data because, differently from static cameras, the robot should be moving among people for an extensive period of time and collect a multitude of views, which can be technically infeasible. Therefore, deep-learning approaches based on large amounts of thermal data are not suitable for our human re-identification system. Moreover, in domestic environments, people can move freely and can be observed by the robot from many different views and distances. Thus, it is essential to implement a robust re-identification system that can cope with these variations and the uncertainty introduced by occlusions and human pose.

Our approach builds a dictionary of thermal features at different views. However, instead of sampling at predefined, we perform an entropy-based sampling that automatically selects the observations providing more information. We then transform each video to a new sequence of dictionary elements (symbols). In this new representation, we use the geometric distribution among symbols as features and train a support vector machine (SVM) classifier. As our approaches performs entropy-based sampling, it uses small amount of data, which fits well with the requirements of service robots.

Although thermal images are widely used in computer vision, especially for face recognition, there is no benchmark dataset for re-identification. To our knowledge, there are two datasets about people walking i) along a corridor [19] and ii) outside a building [22], both recorded by a fixed camera in a surveillance setup, not covering the case of domestic environments from a robot perspective. Thus, we have collected a thermal re-identification dataset, which is publicly available, with the camera mounted on a mobile service robot (see Fig. 1a). We have recorded thermal images of people under various domestic environment cases such as walking, sitting, occlusion from different views.

The contributions of this paper are threefold:

- a new entropy-based sampling to build thermal dictionaries and symbolic representations of humans in thermal images;
- a full software pipeline, implemented in ROS², for thermal-based human re-identification with service robots;
- a new publicly available thermal dataset for human re-identification with such robots.

The reminder of this paper is as follows. Related work on re-identification approaches are presented in Section 2. Section 3 explains the details of our approach and how entropy-based sampling is used to create thermal dictionary models (TDM). The symbolic representation and classification in the TDM domain are described in Section 4. Experimental results with our new public dataset are presented in Section 6. Finally, we conclude this paper in Section 7 discussing achievements and current limitations.

¹ENRICHME: ENabling Robot and assisted living environment for Independent Care and Health Monitoring of the Elderly – <http://www.enrichme.eu>.

²Robot Operating System – <http://www.ros.org>.

2 Related Work

The main goal of re-identification is to establish a consistent labeling of the observed people across multiple cameras or in a single camera in non-contiguous time intervals [2].

The approach of [10] on RGB cameras focuses on an appearance-based method, which extracts the overall chromatic content, spatial arrangement of colors and the presence of recurrent patterns from the different body parts of the person. In [17], the authors propose a deep architecture that automatically learns features for the optimal re-identification. However, the problem of these methods is the use of color, which is not discriminative for long-term applications.

In [1], re-identification is performed on soft biometric traits extracted from skeleton data and geodesic distances extracted from depth data. These features are weighted and used to extract a signature of the person, which is then matched with training data. The methods in [20, 21] tackles the problem applying features based on the extracted skeleton of the person. This is used not only to calculate distances between the joints and their ratios, but also to map the point clouds of the body to a standard pose of the person. This allows to use a point cloud matching technique, typical of object recognition, in which the objects are usually rigid. However, as skeleton data is not robust to body motion and occlusion, these approaches have strong limitations. In addition, point cloud matching has a high computational cost.

In [25], a multi-modal dissimilarity representation is obtained by combining appearance and skeleton data. Similarly, in [24], an ensemble of distance functions, each one learned using a single feature, is built in order to exploit multiple appearance features. While in other works the weights of such functions are pre-defined, in the latter they are learnt by optimizing the evaluation measures. Although these ensembles of state-of-the-art approaches can improve the accuracy of human re-identification, their dependency on color and/or skeletal data pose strong limitations on the type of environment and sensing capabilities of a mobile robot.

In [3], human recognition is performed by fusing a histogram-based human clothes classification, which takes into account the uncertainty of the human position to select relevant image regions, and a simple face

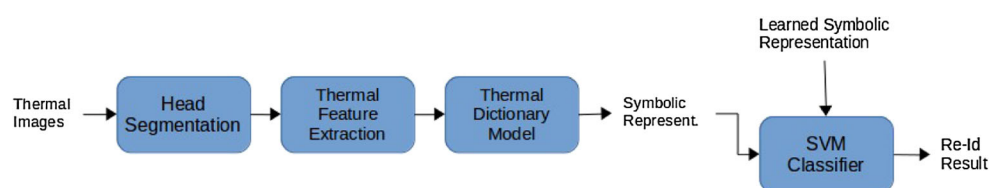
recognition algorithm. Then, the output of the recognizers are integrated with multi-sensor detectors to perform simultaneous tracking and recognition. Wengefeld et al. [30] present a combined system on a mobile robot using both laser and 3d-camera for detection, tracking and visual appearance based re-identification. Similarly, [14] presents a method for person identification and tracking with a mobile robot. The person is recognized using height, gait, and appearance features. The tracking information is also used in [29], where the identification is based on an appearance model, using particle swarm optimization to combine a precise upper body's pose estimation and appearance. In these approaches, re-identification is mainly used to recover human IDs during people tracking. In this case, appearance based features are enough for human re-identification in the short period, but not to identify people in the long term.

In the last decade, thermal images have become increasingly popular to solve standard computer vision problems, in particular for face recognition [7, 11, 12, 31]. In [7], local binary patterns (LBP) are extracted from thermal images. Then, given a feature vector from a test sample, authors use partial least squares-discriminant analysis (PLS-DA) to perform face recognition. Wu et al. [31] presented a convolutional neural network (CNN) architecture that can automatically learn effective features from thermal data and perform face recognition with a softmax classification layer. Although there are many thermal face recognition approaches that achieve good results, they are typically not suitable for domestic service robot applications because they require a clear frontal image of the face, which is often not possible to obtain with a mobile robot.

3 Entropy-Based Sampling

The proposed approach for human re-identification uses images acquired by a thermal camera, shown in Fig. 3a. It performs face segmentation, extracts thermal features and creates thermal dictionary models from training sequence. Symbolic representations of the thermal features are used then to train an SVM classifier. The flow diagram of the system and the respective sub-modules are depicted in Fig. 2.

Fig. 2 The flow diagram of our approach for thermal-based human re-identification



The following subsections explain each part of our approach in detail.

3.1 Head Segmentation and Thermal Feature Extraction

The image acquired from the thermal camera provides the temperature of objects in the field of view of the camera. Since the temperature of humans are within a specified interval, it is possible to segment people in the thermal image by thresholding the temperature data. Face and body provide an important feature to recognize people. However, the temperature data obtained by observing the human body is largely dependent on the type of clothes the person wears (Fig. 3b). Therefore, we focus on the segmentation of the head region only.

We first perform thresholding on the thermal image (Fig. 3b) in the interval $[32\text{ }^{\circ}\text{C} - 39\text{ }^{\circ}\text{C}]$ and obtain a binary image (Fig. 3c). Then, we apply connected component analysis on the binary image. We filter the components based on area and width, by keeping the ones that occupies an area and that has a width bigger than pre-defined values. Among the remaining components, we select the region of the binary image with smallest width (Fig. 3). After the head region on the thermal image is segmented, we extract features from it. The temperature

data of the head region (i.e., whole 3D head, not just 2D face) provide important information. We therefore calculate the temperature histogram of the current head region (Fig. 3e), which is normalized to obtain the distribution of the temperatures (Fig. 3f). The concatenation of temperature histograms from different points of view will provide the temperature characteristic of the person's head. In [27], it is noted that the temperature of the skin surface varies with the environmental temperature, the body temperature, with the conditions of the skin and the structures beneath it. However, the head is one of the regions on the human body where the skin temperature remains more or less constant despite temperature changes in the environment [23]. Our hypothesis is that the head temperature distribution can be used to distinguish people identities. This is verified experimentally in Section 6.

3.2 Thermal Dictionary Models

In real-world scenarios, where people freely move in the environment, service robots require a view-independent re-identification approach. Considering single-shot re-identification, most of the view-independent methods are based on full person models built from various views [6, 33]. Although they can typically achieve very good results, they also have a high computational cost. In addition, finding

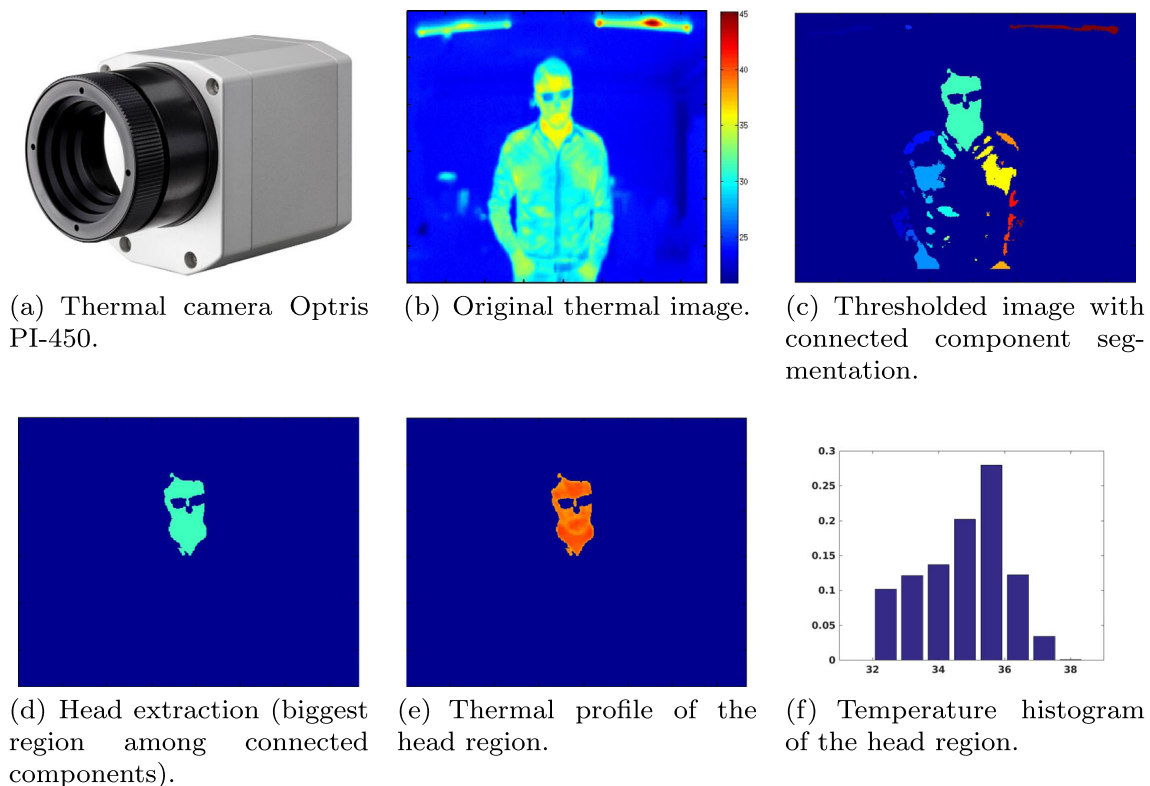
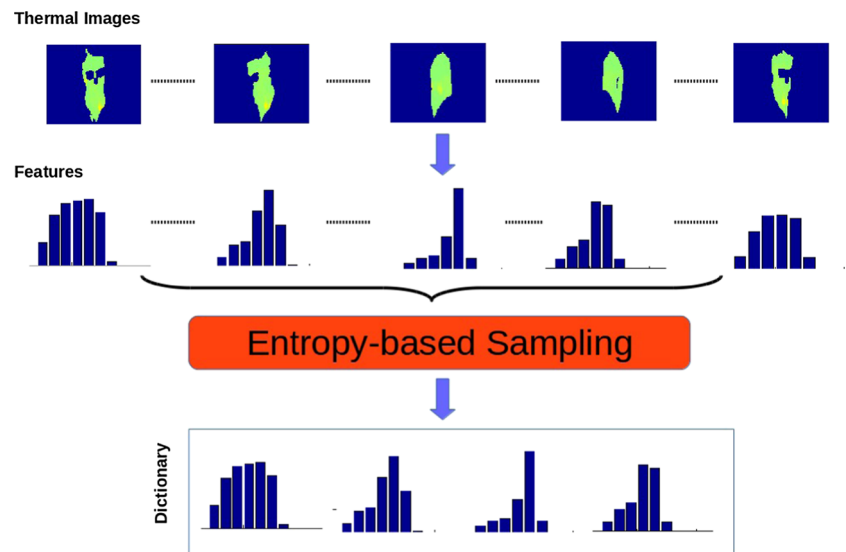


Fig. 3 Head segmentation and thermal feature extraction

Fig. 4 An example of entropy-based sampling to find the most informative features representing a person at different views



the optimal quantization of the possible viewing angles is not easy. In light of these considerations, we exploit people motion and model the related change of thermal features, instead of learning models at pre-defined angles.

A model representing the person from different point of view can be embedded in a dictionary of thermal features. Assuming we have thermal data of a person turning around, we can extract a sequence of features obtained at different angles (Fig. 4). However, choosing the sampling angles is not easy. The representation could be too coarse or too fine, depending on the pre-defined angle intervals. In our approach, instead, we let the data determine which features are worth to be kept by performing an entropy-based sampling on the sequence of features. The features that provide sufficient information gain are included into the dictionary.

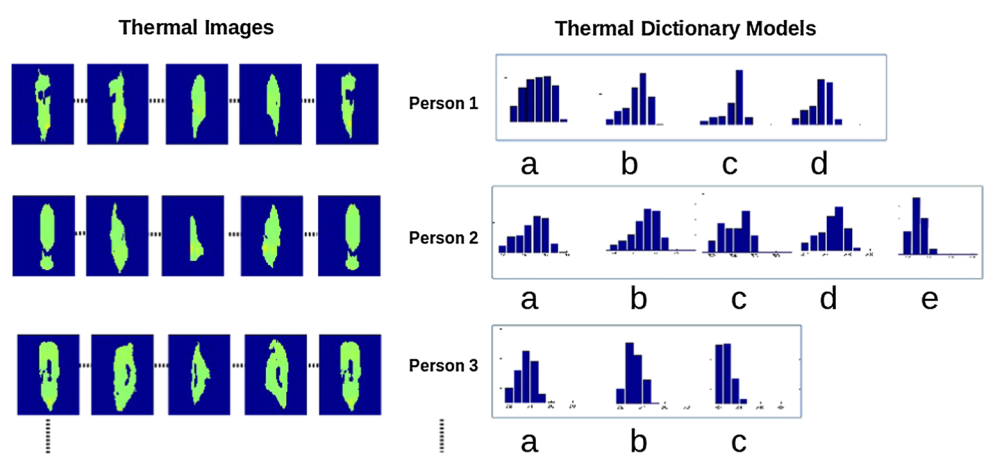
In information theory, the information gain (relative entropy) is a measure of the difference between two probability distributions, which can be measured by the Kullback-

Leibler (KL) divergence [15]. We determine the information gain between features using the latter, calculated as follows:

$$D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (1)$$

where P and Q are features extracted from the thermal data. Using Eq. 1, we perform an entropy-based sampling as follows: first, we calculate the KL divergence between each element in the dictionary model and a new thermal feature; then, if the information gain is bigger than a pre-defined threshold, we include the new feature to the dictionary model. The procedure is summarized in Algorithm 1. Then, we convert the sampled features into symbols and include them in a *thermal dictionary model* (TDM) (Fig. 5). As TDMs are individually processed and generated for each person, we associate histograms from different people with same symbols.

Fig. 5 Using entropy-based sampling, we create a thermal dictionary model of each person and represent it in symbols



In our approach, we do not assume a pre-defined number of samples or any particular (fixed) orientation of the user. Thanks to our entropy-based sampling scheme, the system automatically selects the most informative observations. During training, the user is only asked to turn around in front of the robot, so the system can measure the information gain and automatically generate the TDMs. These are used for the classification step of the re-identification, as explained in the next section.

Algorithm 1 Thermal dictionary models are obtained by entropy-based sampling based on the Kullback-Leibler (KL) divergence [15].

```

TDM: dictionary model of a subject;
foreach  $\tilde{F}$ : new thermal feature observed in training set;
do
  foreach  $F_i \in TDM$  do
    Calculate the information gain of the new thermal feature;
     $Dist_i = D_{KL}(\tilde{F} \| F_i)$ ;
    if any of  $Dist_i > Thr$  then
      Insert the new thermal feature into dictionary model;
       $\tilde{F} \rightarrow TDM$ 
    end
  end
end

```

4 Classification in the TDM Domain

4.1 Symbolic Representation

For each person, we extract a TDM using a training sequence, i.e., T_c , for $1 \leq c \leq C$ where C is the number of people (classes). Then, we obtain a symbolic representation

of each test sequence by converting the thermal features into symbols using the TDM of each person. For each feature extracted, we find the most similar element in the TDM and assign its symbol to the feature. Here, we use the KL divergence (1) as a measure to evaluate the similarity between features and TDM elements. We calculate the KL divergence between a feature and each TDM element and assign the symbol of the most similar element. As a result of this operation, we obtain a symbol S for each TDM:

$$S = \text{symbol}(T^m) \quad (2)$$

$$m = \arg \min_i D_{KL}(F \| T^i)$$

where T^i represents the i^{th} element of the TDM, F is the extracted feature, and m is the index of the most similar element.

An important aspect of the symbolic representation lies on taking into account how similar a feature is to a dictionary element. S represents the most similar element of each feature, but it does not contain this information. Thus, we extend the symbolic representation by including the similarity measure between features and dictionary elements (3).

$$M = D_{KL}(F \| T^m) \quad (3)$$

In conclusion, we obtain the following combined representation for each thermal feature:

$$F \rightarrow \Sigma = \{S, M\} \quad (4)$$

4.2 Classification

Σ^c provides a representation of feature vector F in the domain of samples from class c . This representation contains answers to the following questions about the feature vector: i) what is the most similar dictionary element in class c and ii) how similar are they? Geometrically, the representation is shown in Fig. 6a.

Fig. 6 Symbolic representation of a feature vector in class 1 (i.e., person 1) (a) and whole training space (b). Light and dark colored stars represents all the test samples and the dictionary elements of each class obtained by entropy-based sampling (e.g., T_1), respectively

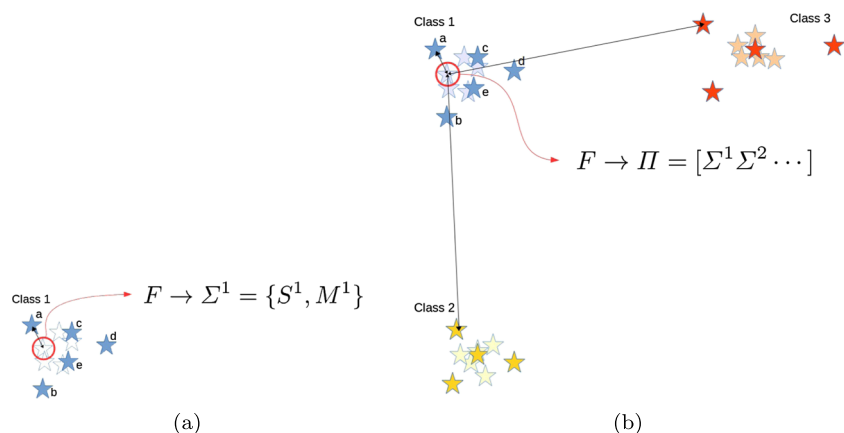
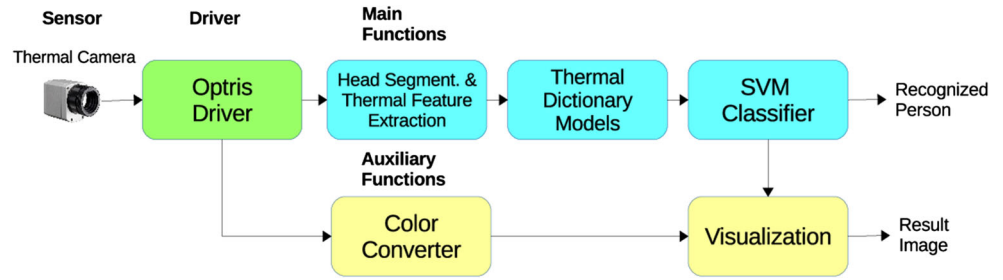


Fig. 7 The ROS software pipeline for thermal-based human re-identification



If we concatenate the symbols computed for each class into a single vector, we obtain a new feature vector with respect to the whole training space (Fig. 6b):

$$\Pi = [\Sigma^1 \dots \Sigma^C] \quad (5)$$

With this representation, we can encode any new feature vector in the (previously trained) feature space. We assume that features from the same class (i.e., person) will have similar representations, so we can match the representation of a test sample to the representation of training samples from the same class.

As a result of the features extracted by Eq. 5, we obtain a high dimensional data representation for the next classification stage. We train an SVM classifier [9], which is proven to work very well for high dimensional data [13]. We use the similarity measure to train the SVM classifier:

$$\begin{aligned} X &= [M^1 \dots M^C] \\ Y &\in \{1, \dots, C\} \end{aligned} \quad (6)$$

5 Software Pipeline and Implementation

The full software pipeline of the proposed human re-identification approach is implemented as a ROS node³ for applications with domestic mobile robots. This ROS node assumes that the SVM classifier is already trained and the corresponding models for thermal dictionary and classifier are present. The software pipeline is illustrated in Fig. 7. The blue and yellow boxes present the main and auxiliary functions implemented in the software, respectively. The green box represents the ROS driver of the Optris thermal camera⁴. The following subsections explain the details of the pipeline.

5.1 Thermal Image Acquisition

In this paper, the thermal images are acquired using the thermal camera Optris PI-450 (Fig. 3a). This thermal camera offers a temperature range of -20°C up to 900°C , with a thermal sensitivity of 0.04°C . It has a spectral range of 7.5 to $15\ \mu\text{m}$. The image resolution is 382×288 and the maximum frame rate is $80\ \text{fps}$. Our model is equipped with a 15mm lens, which provides a field of view (FOV) of $38^\circ \times 29^\circ$. The ROS driver of the Optris thermal camera provides in two different formats: i) raw image data in unsigned short format, ii) color thermal image in (BGR8) format. Using the ROS driver of Optris thermal camera, both raw and color thermal images are acquired and made directly available to our ROS software module for thermal re-identification.

$^\circ\text{C}$, with a thermal sensitivity of 0.04°C . It has a spectral range of 7.5 to $15\ \mu\text{m}$. The image resolution is 382×288 and the maximum frame rate is $80\ \text{fps}$. Our model is equipped with a 15mm lens, which provides a field of view (FOV) of $38^\circ \times 29^\circ$. The ROS driver of the Optris thermal camera provides in two different formats: i) raw image data in unsigned short format, ii) color thermal image in (BGR8) format. Using the ROS driver of Optris thermal camera, both raw and color thermal images are acquired and made directly available to our ROS software module for thermal re-identification.

5.2 Software Implementation

The thermal re-identification software is encapsulated into a ROS package, which can be very easily installed thanks to catkin-compatibility⁵. The ROS package developed for re-identification includes a one-click roslaunch⁶ file with a YAML⁷ file containing the parameters for re-identification shown in Table 1.

This ROS node subscribes to both raw and color thermal images published by the Optris driver and performs the following operations. First, following the PI imager library⁸, raw image data (*data*) are converted to temperature values in floating point format (*t*) as follows:

$$t = (data - 1000)/10.0 \quad (7)$$

Then, as described in Section 3, the ROS node performs thermal feature extraction and create the symbolic representation of a new thermal feature. Finally, the SVM classifier is used to predict the label of the new feature. The SVM classifier is implemented by using the libSVM library [4]. This ROS node assumes that the model files for thermal dictionary and classifier are present under the “config” directory of the ROS package.

The name of the recognized person is published as a ROS topic together with a confidence level. In addition, the result

³https://github.com/scosar/Thermal_Reidentification

⁴http://wiki.ros.org/optris_drivers

⁵<http://wiki.ros.org/catkin>

⁶<http://wiki.ros.org/roslaunch>

⁷<http://yaml.org/>

⁸<http://documentation.evocortex.com/libirimag2/html/index.html>

Table 1 Description of the parameters of our thermal re-identification software provided with their default values

Name	Description	Default value	Unit
<i>temp_thr_min</i>	Lower value for thresholding thermal images	32	°C
<i>temp_thr_max</i>	Higher value for thresholding thermal images	39	°C
<i>component_thr</i>	Minimum value for the area of the head region	400	pixels
<i>component_width</i>	Minimum value for the width of the bounding box of the head region	110	pixels

of the re-identification is visualized on a color converted thermal image and published as a separate ROS topic.

5.3 Training

The training of the SVM classifier is performed by an offline procedure using a separate software. We set the parameters of libSVM [4] to perform multi-class support vector classification. Linear, polynomial, sigmoid and radial basis functions were tried as kernels. The radial basis function (RBF) was empirically selected to provide the best classification results. The γ of the RBF is set to $1/120$, where 120 is the size of our feature vector. Different values of the cost (C) and epsilon ϵ have been tried. In particular, we have tested the following ranges: $C =$

$\{10^k, k = -1, 0, 1, 2, 3\}$, $\epsilon = \{0.01, 0.001, 0.0001\}$. Based on these classification tests, we chose the C and ϵ that gave the best results, i.e. 1000 and 0.01, respectively. To reduce the training time, we also enabled the `shrinking` parameter of libSVM and set the kernel cache to 4,000 MB.

6 Experimental Results

We evaluated the performance of our approach under various real-world conditions such as walking, sitting, and occlusion of face. As one of the contributions of this paper, we have recorded a novel thermal dataset for human re-identification. The details of this dataset together with the obtained results are presented in the following subsections.

Fig. 8 The dataset consists of four parts: people standing still and turning around (a–b), people walking freely (c), people sitting on a sofa recorded at 3 different views (d–f), and people occluding parts of their face while wearing hat, glasses, and scarf (g–i)

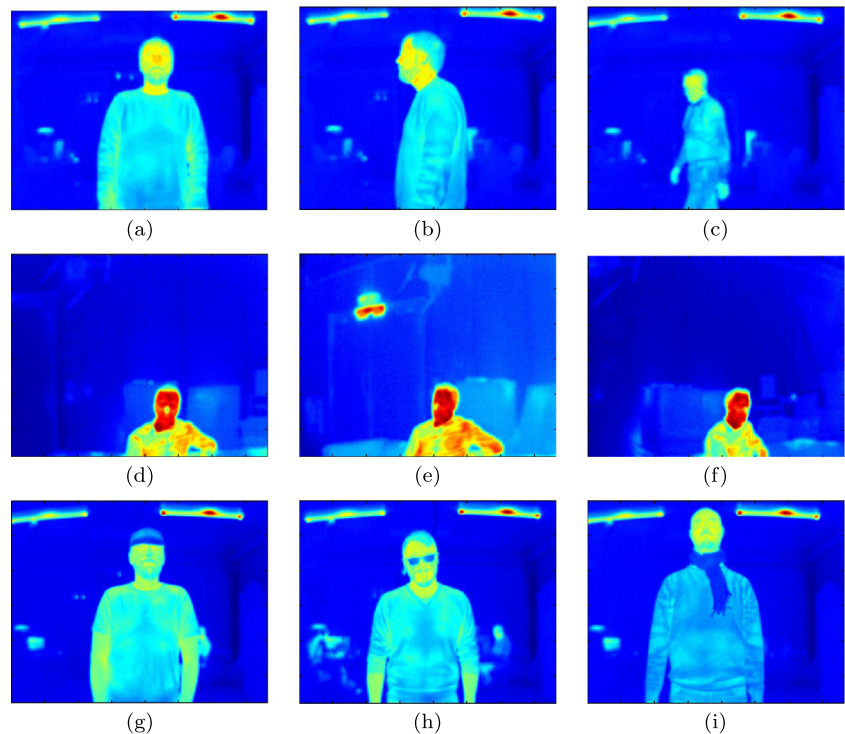


Table 2 The details of the sequences in the thermal re-identification dataset

Sequence	Number of Instances	Duration (per instance)	Number of Images
Turn Around	6	15s	900
Occlusion	4	15s	600
Walking	2	30s	600
Sitting	2	15s	300
Total	210	1125s	36000

The sequence details are presented as per person

6.1 Thermal Re-identification Dataset

We have recorded a publicly available thermal dataset for human re-identification⁹. The dataset was recorded in a laboratory environment using an Optris PI-450 thermal camera mounted on a Kompaï robot. Thermal images were recorded with a resolution of 382×288 at 10 fps. Our dataset covers different challenges in real-world scenarios, such as observing people from different points of views, while walking and sitting. It also includes people wearing accessories such as hat, glasses, and scarf that occlude part of the face. The dataset is available as ROSbag files and it can be directly used in a ROS environment.

In particular, the dataset consists of sequences of four categories: 1) person turning around on the spot, 2) person walking, 3) person sitting on a sofa, and 4) person turning around while wearing hat, glasses, or a scarf. In total, 15 people were recorded in the dataset. Sample images from the dataset are depicted in Fig. 8.

For the turn around and occlusion sequences (categories 1 and 4), the participants were asked to turn to four different directions (frontal, left side, back, right side) and remain still for about 3 seconds. In the occlusion sequences, they were also asked to wear accessories occluding different parts of the face. We asked people to repeat the sequences 4 times and 6 times in the occlusion and turn around sequences, respectively. On average, each sequence lasted about 15 seconds, resulting in 600 and 900 total thermal images per person in occlusion and turn around sequences, respectively. The walking sequences (category 2) lasted about 30 seconds. They contain images of non-frontal face views of people walking freely in front of the robot. The sitting sequence (category 3) contains images acquired while the participants were sitting on a sofa at three different distances (2m, 3.5m, and 5m) for about 5 seconds. For the walking and sitting sequences, we asked people to repeat the same activity twice, resulting in 600 and 300 thermal images per person, respectively. In total, our dataset contains around

2,400 thermal images per person, resulting 36K images in total (Table 2).

Thermal image datasets are relatively new in computer vision and robotics communities. Most of the existing datasets are for face recognition and include only frontal images [7]. To our knowledge, there are only two thermal datasets for human re-identification that do not focus on faces [19, 22]. However, these datasets only contain thermal images of people walking along a corridor or in front of a building. Therefore, they are not suitable to represent real-world situations typical of domestic environments (e.g., sitting). To the best of our knowledge, ours is the first thermal dataset recorded using a robot in challenging real-world scenarios.

6.2 Experimental Setup

In our experiments, we used the sequences in which people turn around on the spot for learning the thermal dictionary models, and the rest for testing. We took 2/3 of the turn around set for training. The rest is included in the testing set. Hence, on average, 600 thermal images per person were used for training.

For the thermal features, we calculated histograms (10 bins) of the head region in the same temperature interval used for thresholding, i.e. $[32^\circ\text{C} - 39^\circ\text{C}]$. The number of bins was selected empirically from real tests. With a higher number of bins, the histograms become sparse, and the TDMs very large, decreasing the re-identification performance. If we select a smaller number of bins, the histograms become flat and look mostly alike, generating very few TDMs, which are not enough to distinguish people.

6.3 Results

We evaluated the system on single-frames comparing the recognized class of each frame in a test sequence with the ground truth. We compared our approach (“Symbolic Rep.”) to a standard SVM classifier using the whole training set without entropy sampling (“Whole Training Set”). In order to analyse the advantages of the symbolic representation, we also compared our approach to an SVM classifier that

⁹<https://lcas.lincoln.ac.uk/wp/research/data-sets-software/l-cas-rgb-d-t-re-identification-dataset/>

Table 3 The re-identification accuracy rates of our approach (Symbolic Rep.), SVM with whole training set and entropy-based sampling for turn around sequences

Methods	Turn around (%)				
	Recall	Precision	Accuracy	F_1 Score	nAUC
Whole Training Set	75.1	76.2	76.8	75.7	97.5
Entropy-based Samp.	76.3	77.0	77.4	76.6	97.6
Symbolic Rep.	87.4	87.2	87.4	87.3	98.2

Bold values represent the best accuracy rates

uses plain TDMs (without symbolic representation) after entropy-based training (“Entropy-based Samp.”).

We calculated recall, precision, accuracy and F_1 score for every subject individually, averaging the results across all the test frames. We also computed the Cumulative Matching Characteristic (CMC) curve, which is commonly used for evaluating re-identification methods [28]. For every $k = \{1 \cdots N_{train}\}$, where N_{train} is the number of training subjects, the CMC expresses the average person recognition rate computed when the correct person appears among the k best classification scores (rank- k). A standard way to evaluate CMC is to calculate the rank-1 recognition rate and the normalized Area Under Curve (nAUC), which is the integral of the CMC.

6.3.1 Turn Around Sequences

First, we evaluated the system on turn around sequences. As the type of motion in the test set is the similar to training, the re-identification problem is relatively easy and we would expect very good results.

This is confirmed indeed by Table 3. It can be seen that our symbolic representation achieves the best results with over 85% recall, precision and accuracy. This is much

better than the 75% obtained with the other approaches. We should also note that our symbolic representation achieves 10% better performance in average compared to the SVM classifier without symbolic representation (“Entropy-based Samp.”). Figure 9 shows the CMC curve of all the approaches for the turn around sequences. Again, this shows that our symbolic representation outperforms the other approaches.

6.3.2 Occlusion Sequences

The results for the occlusion scenario are presented in Table 4. This experiment was designed to understand the effect of occlusion cases that usually happen in real-world scenarios wearing accessories such as glasses, hat, and scarf. The results show a decrease in performance compared to the turn around case, which was expected due to the more challenging nature of the experiment. However, our approach still achieves the highest re-identification rates among all methods. We can also see that our symbolic representation achieves 5% better performance in average compared to the classifier without symbolic representation. This is also confirmed by the CMC curve in Fig. 10. Although the results look similar in higher ranks, we can

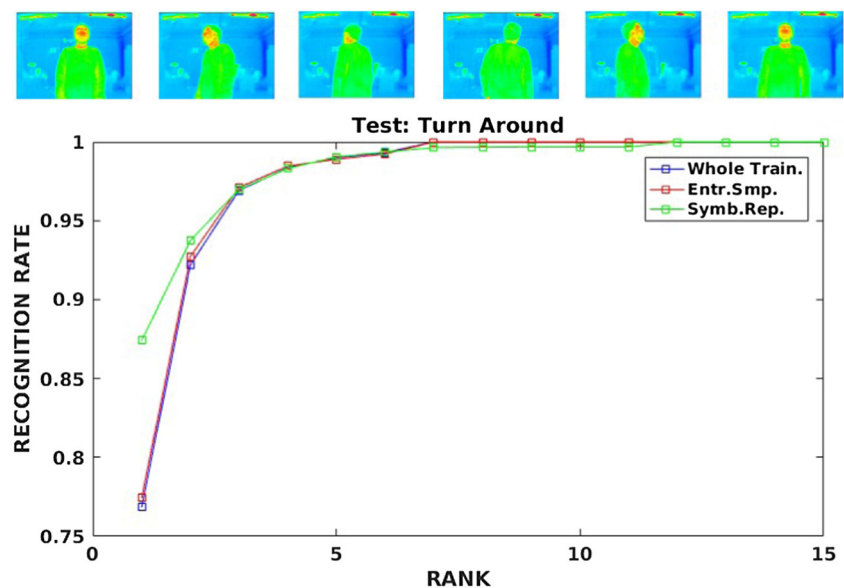
Fig. 9 The Cumulative Matching Characteristic (CMC) curve of all approaches for turn around sequence

Table 4 The re-identification accuracy rates of our approach (Symbolic Rep.), SVM with whole training set and entropy-based sampling for occlusion sequences

Methods	Occlusion (%)				
	Recall	Precision	Accuracy	F_1 Score	nAUC
Whole Training Set	51.0	54.5	53.2	52.7	91.0
Entropy-based Samp.	53.4	56.4	55.1	54.8	91.3
Symbolic Rep.	59.0	61.7	59.9	60.3	91.1

Bold values represent the best accuracy rates

clearly see that our approach obtains the highest recognition rate in lower ranks, especially in rank-1.

6.3.3 Walking Sequences

Table 5 presents the re-identification results for the walking scenario. This includes people walking freely observed by various angles and distances very different from the training set. Thus, the complexity of this scenario is higher than previous cases, as can be observed by the performance drop for all the methods. Nevertheless, our approach achieves again the best performance, in particular thanks to the improvement introduced by the entropy-based sampling stage. This can also be observed from the CMC curves in Fig. 11, showing that our approach obtains the highest recognition rate in most of the lower ranks.

6.3.4 Sitting Sequences

Finally, in the sitting sequences, we evaluated the performance under different human poses. This scenario is to understand the effects of observation distance individually.

The results are presented in Table 6. We can say that our approach is not much affected by distance. It clearly outperforms other approaches. Compared to the classifier without symbolic representation, we can see that our final classification achieves 5% better performance in average. Again, when we look at CMC curve in Fig. 12, we can see that our approach achieves the best re-identification rates almost in all ranks.

6.3.5 Overall Performance

For a comparison of overall re-identification performance, we have measured the accuracy of all approaches in a big testing set that consists of all testing sequences. Table 7 presents the overall results and Fig. 13 displays the overall CMC curve. Once again, we can clearly see that our approach outperforms the others with a high margin. We can also see that our symbolic representation achieves 10% better accuracy in average compared to the SVM classifier using only plain TDMs (without symbolic representation). This shows the superiority of the symbolic approach.

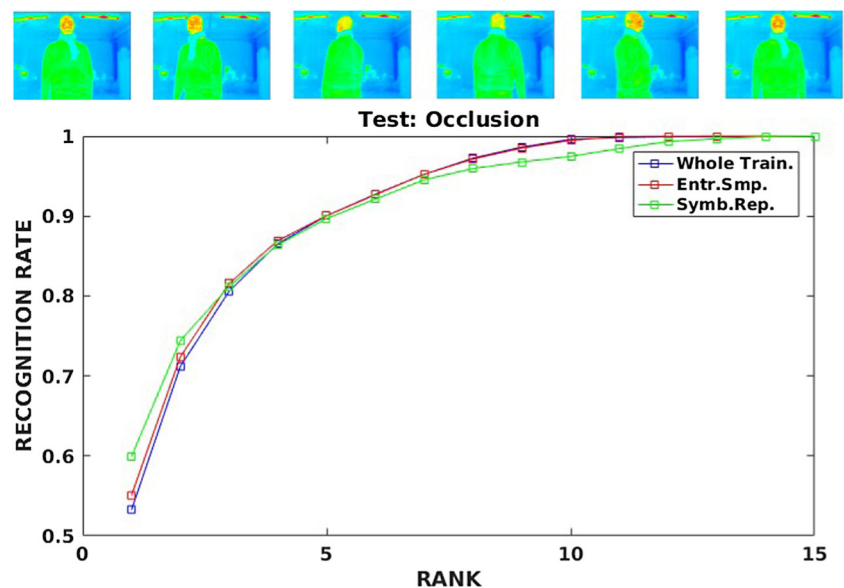
Fig. 10 The Cumulative Matching Characteristic (CMC) curve of all approaches for occlusion sequence

Table 5 The re-identification accuracy rates of our approach (Symbolic Rep.), SVM with whole training set and entropy-based sampling for walking sequences

Methods	Walking (%)				
	Recall	Precision	Accuracy	F_1 Score	nAUC
Whole Training Set	24.7	27.0	23.9	25.8	78.7
Entropy-based Samp.	30.7	39.4	29.6	34.5	80.3
Symbolic Rep.	31.0	40.3	27.8	35.0	80.5

Bold values represent the best accuracy rates

6.4 Discrimination and Rejection

In this subsection, we present the experiments that are performed to evaluate the discrimination and rejection properties of our approach.

To have a better understanding of the classification results, we calculated the confusion matrix for the turn around sequence in Fig. 14. We can see that, except for a couple of people (person 10 and 14), our approach achieves at least 80% recognition accuracy. This also proves that our symbolic representation, even if based on simple temperature histograms, enables a powerful and discriminative re-identification of humans with a robot thermal camera.

We have also tested the rejection property of our approach by analysing the confidence level of the samples that were correctly classified (true positive). In Fig. 15, we present the histogram of all the confidence levels in the testing sequences. We can see that more than 90% of the true positives have a confidence level greater than 60%, proving that our approach has the ability to reject unknown people in most of the cases.

6.5 Effects of Head Segmentation Performance

The head segmentation process is the first step of our algorithm, and its failure may compromise the re-identification performance. To understand the effect on our system, we simulated several levels of failure in the head segmentation by artificially occluding the binary image from four directions (Fig. 16b-c), replicating potential problems due to poor thresholding or connected component analysis. In particular, we tested the classification performance by occluding 10% to 90% (at 10% intervals) of the binary image in the turn around sequence.

Figure 17 shows the recognition rate of our approach with various levels of occlusion. Similar to the previous occlusion experiments (Section 6.3.2), we see that there is a decrease in the recognition rate as the face gets more and more occluded. However, we can also see that our approach still works in challenging cases, achieving 60% recognition rate with 30% occlusion. Notice that the recognition performances with the occlusions on the y-axis (top-to-bottom and bottom-to-top) are slightly worse than

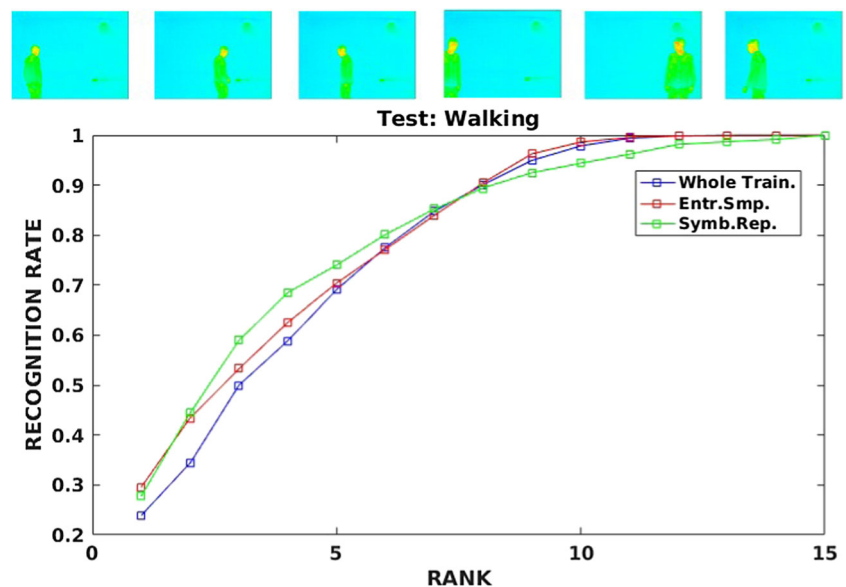
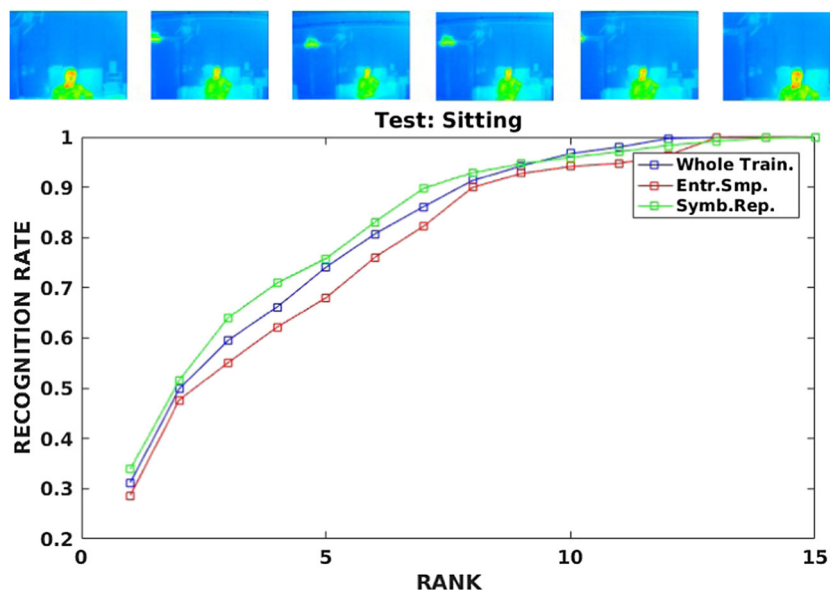
Fig. 11 The Cumulative Matching Characteristic (CMC) curve of all approaches for walking sequence

Table 6 The re-identification accuracy rates of our approach (Symbolic Rep.), SVM with whole training set and entropy-based sampling for sitting sequences

Methods	Sitting(%)				
	Recall	Precision	Accuracy	F_1 Score	nAUC
Whole Training Set	33.4	30.9	31.2	32.1	81.9
Entropy-based Samp.	30.5	27.9	28.7	29.1	79.2
Symbolic Rep.	34.7	36.7	33.8	35.7	83.1

Bold values represent the best accuracy rates

Fig. 12 The Cumulative Matching Characteristic (CMC) curve of all approaches for sitting sequences**Table 7** The overall re-identification accuracy rates of our approach (Symbolic Rep.), SVM with whole training set and entropy-based sampling

Methods	Overall(%)				
	Recall	Precision	Accuracy	F_1 Score	nAUC
Whole Training Set	46.5	48.7	47.6	47.6	88.3
Entropy-based Samp.	47.9	49.7	48.7	48.8	88.1
Symbolic Rep.	56.4	59.0	57.6	57.7	90.3

Bold values represent the best accuracy rates

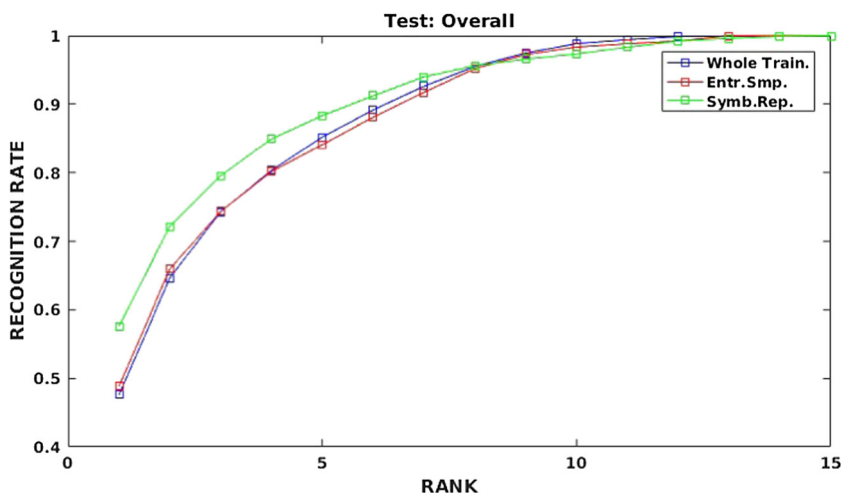
Fig. 13 The overall Cumulative Matching Characteristic (CMC) curve of all approaches

Fig. 14 The confusion matrix of our re-identification system for the turn around sequence

		Test: Turn Around														
True Class	Prs 1	96.7%	0.0%	0.0%	0.0%	9.9%	0.6%	0.0%	5.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Prs 2	0.0%	90.0%	0.0%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	19.5%	5.6%	0.0%	0.0%	1.9%	0.8%
	Prs 3	0.0%	3.6%	89.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	0.0%	0.0%	12.5%	0.0%
	Prs 4	0.0%	0.0%	0.0%	94.4%	0.0%	0.0%	0.0%	0.0%	0.0%	3.9%	2.8%	0.0%	0.0%	11.8%	0.3%
	Prs 5	0.0%	0.0%	0.0%	0.0%	90.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Prs 6	0.0%	0.0%	0.0%	0.0%	0.0%	87.1%	0.0%	0.0%	0.0%	0.0%	0.0%	5.4%	0.0%	0.0%	0.0%
	Prs 7	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	100%	0.0%	0.0%	0.0%	0.0%	3.1%	1.7%	0.0%	0.0%
	Prs 8	2.8%	0.4%	0.0%	2.3%	0.0%	0.0%	0.0%	84.5%	0.0%	1.5%	1.1%	0.0%	0.0%	0.0%	0.0%
	Prs 9	0.0%	0.0%	9.6%	0.0%	0.0%	0.0%	0.0%	0.0%	98.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	Prs 10	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	65.9%	6.7%	3.7%	0.0%	4.4%	5.5%
	Prs 11	0.0%	0.8%	0.9%	0.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	81.3%	0.0%	0.0%	0.6%	0.0%
	Prs 12	0.0%	0.0%	0.0%	0.0%	0.0%	11.4%	0.0%	0.0%	0.0%	0.0%	0.0%	85.1%	1.4%	0.0%	0.0%
	Prs 13	0.0%	1.2%	0.0%	0.0%	0.0%	0.6%	0.0%	0.0%	1.7%	0.0%	0.0%	2.6%	96.9%	3.7%	0.0%
	Prs 14	0.0%	2.8%	0.0%	1.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.3%	2.2%	0.0%	0.0%	62.0%	4.4%
	Prs 15	0.5%	1.2%	0.0%	0.0%	0.0%	0.0%	0.0%	10.0%	0.0%	9.0%	0.0%	0.0%	0.0%	3.1%	89.1%
			Prs 1	Prs 2	Prs 3	Prs 4	Prs 5	Prs 6	Prs 7	Prs 8	Prs 9	Prs 10	Prs 11	Prs 12	Prs 13	Prs 14
		Predicted Class														

on the x-axis, showing a higher dependency of our approach on the temperature at the top and bottom of the head.

An error on the measurement of the temperature from thermal images can negatively affect the head segmentation step. To evaluate the effect of this error, we tested our

approach with several levels of temperature error. We simulate this by randomly removing some pixels from the binary image (Fig. 16a). In particular, we tested the classification performance on the turn around sequence by removing 10% to 90% (at 10% intervals) of the binary image.

Fig. 15 The histogram of confidence levels of our approach for the true positive samples in the testing sequences. 90% of the true positives have a confidence level greater than 60% (marked by a red line)

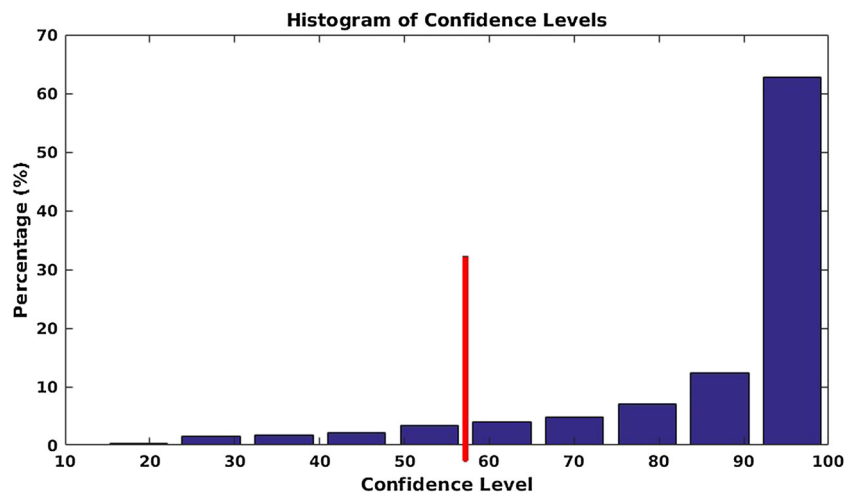
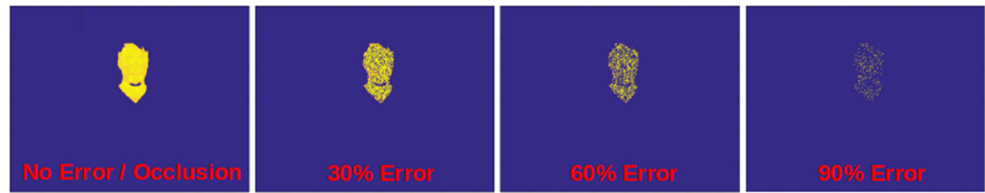
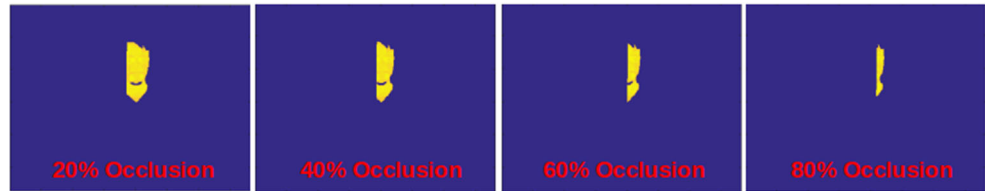


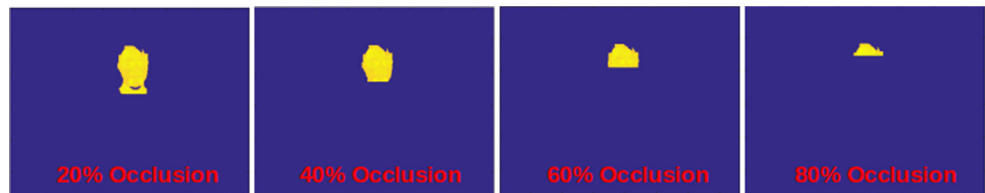
Fig. 16 The effects of head segmentation performance is tested by applying several levels of occlusion and error on the binary image, some examples of which are shown here



(a) Example of binary image with several levels of temperature error.



(b) Example of binary image with several levels of occlusion from left to right.



(c) Example of binary image with several levels of occlusion from bottom to top.

Figure 18 shows the recognition rate of our approach with various levels of temperature error. Our system can still achieve a high recognition rate of 77.8% even with a 50% temperature error, which is higher than the recognition rate achieved by other (non-symbolic) methods without temperature noise (Table 3). This proves that our approach is also robust to this type of errors in the head segmentation step.

6.6 Experiments on a Mobile Robot

We further evaluated our approach on a different platform, a TIAGo mobile robot (Fig. 19a), also used in the ENRICHME project. The system works in real-time on the robot using our ROS software implementation (Section 5). An Optris PI-450 thermal camera was mounted on the robot's head, slightly higher than the previous case. We

Fig. 17 Effects of head segmentation performance

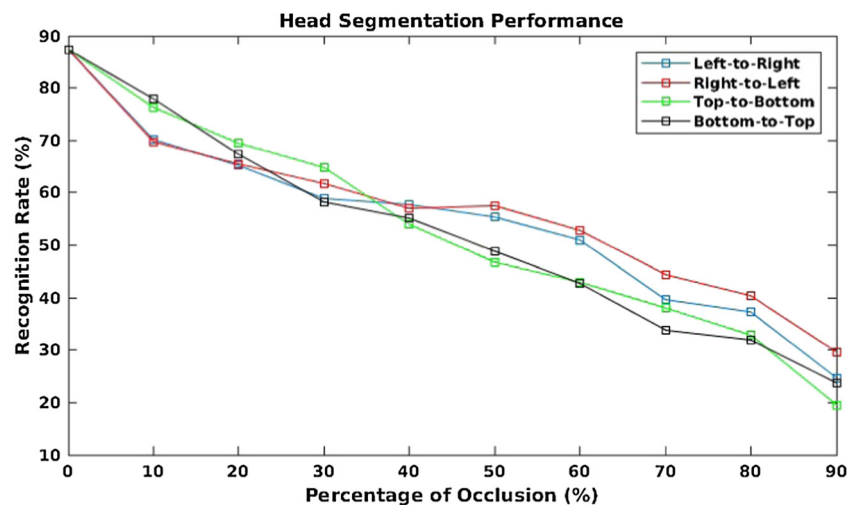
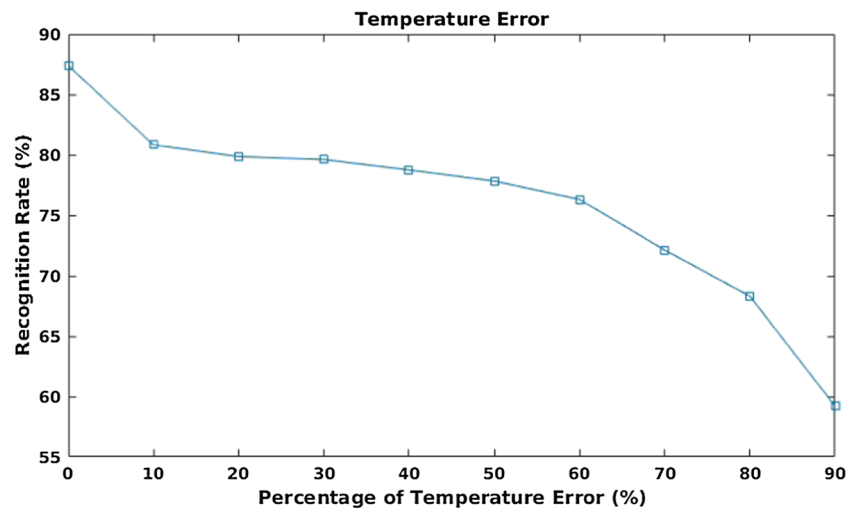


Fig. 18 Effects of temperature error



augmented the training dataset with new data of two subjects, testing the online re-identification system while they were walking and sitting on a chair. A video of these tests is available on the dataset's website.

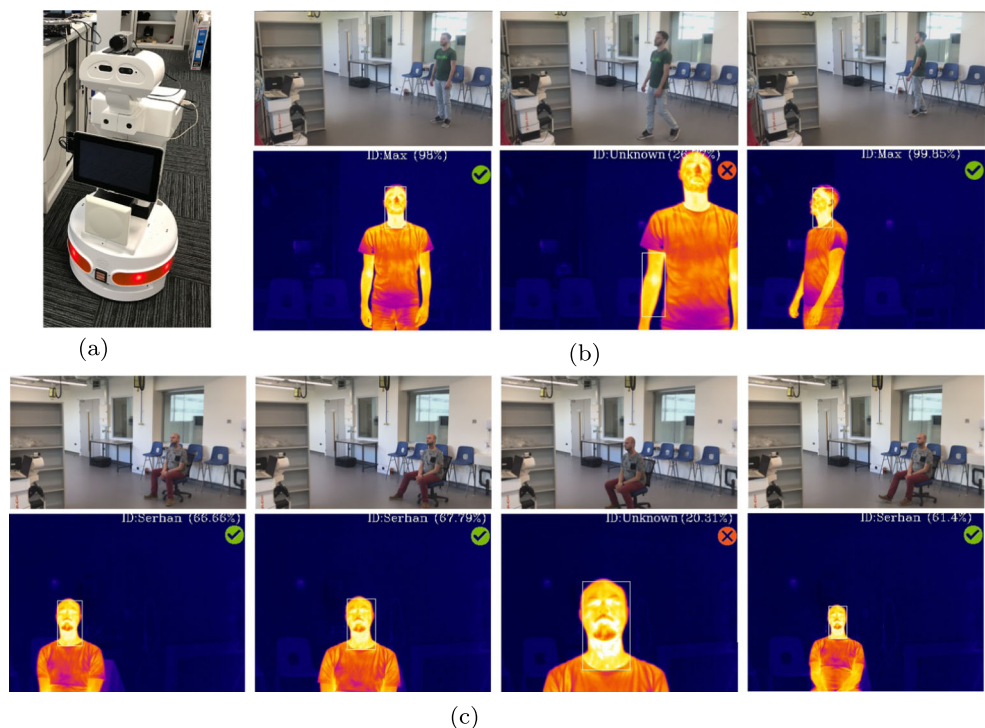
Figure 19 illustrates some outputs of the ROS re-identification software. The images show the segmented head with a white bounding-box and the recognized person together with the confidence value. Following our analysis on confidence level (Section 6.4), we set the rejection threshold on confidence level to 60%. It can be seen that there are some failures, mostly because of wrongly detected head regions (Fig. 19b). We can also notice that the confidence level of the classifier drops when the person is

too close or too far from the robot (Fig. 19c). This is due to the out-of-focus thermal images affecting the extracted features. Nevertheless, we see that in general the proposed approach can recognize the person with high confidence and, importantly, that it works successfully on a real service robot.

7 Conclusion

This paper presented a new re-identification system for service robot applications using thermal images. A view-independent approach, using entropy-based sampling and a

Fig. 19 Some examples of the experiments on a TIAGo mobile robot (a) with two subjects walking (b) and sitting (c). The red marks indicate cases of unsuccessful re-identification



symbolic representation, has been described. The method is suitable for mobile robots monitoring and assisting elderly people at home, in particular to distinguish the actual user in case of two or more occupants. Our solution requires a relatively small amount of training data, which is an advantage in many real-world applications. To achieve this, we extracted a thermal dictionary model of the person sampling over a single rotation sequence of the head. Then, we transform each thermal frames to a new set of dictionary elements (symbols). In this new symbolic representation, we exploit the geometric distributions of its symbols as features for classification. The proposed approach was evaluated under various real-world conditions, including people walking, sitting, and under occlusion. Both quantitative and qualitative results were presented on a new dataset and a real mobile robot, respectively. Despite some limitations in case of walking or sitting people, the experimental results showed the good performance of our re-identification system in several challenging situations and proven that it can be used for companion robots assisting elderly in daily life. Future work will consider temporal models and multi-sensor extensions of our solution to improve the robustness of the re-identification in case of different human poses and motion behaviors. We will also look into on-line learning approaches [18, 32] to incrementally improve the re-identification performance over time as the robot keeps track of people and collects more and more data about the target users.

Acknowledgements This work was supported by the EU H2020 project “ENRICHME” (grant agreement nr. 643691).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Barbosa, I.B., Cristani, M., Del Bue, A., Bazzani, L., Murino, V.: Re-Identification with Rgb-D Sensors. In: European Conference on Computer Vision, pp. 433–442. Springer (2012)
- Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **32**(4), 270–286 (2014). <https://doi.org/10.1016/j.imavis.2014.02.001>
- Bellotto, N., Hu, H.: A bank of unscented kalman filters for multimodal human perception with mobile service robots. *Int. J. Soc. Robot.* **2**(2), 121–136 (2010). <https://doi.org/10.1007/s12369-010-0047-x>
- Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011). Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- Chen, D., Yuan, Z., Hua, G., Zheng, N., Wang, J.: Similarity Learning on an Explicit Polynomial Kernel Feature Map for Person Re-Identification. In: IEEE CVPR, pp. 1565–1573. <https://doi.org/10.1109/CVPR.2015.7298764> (2015)
- Cho, Y.J., Yoon, K.J.: Improving Person Re-Identification via Pose-Aware Multi-Shot Matching. In: 2016 IEEE CVPR, pp. 1354–1362. <https://doi.org/10.1109/CVPR.2016.151> (2016)
- Choi, J., Hu, S., Young, S.S., Davis, L.S.: Thermal to Visible Face Recognition. In: SPIE DSS-DS107: Biometric Technology for Human Identification IX (2012)
- Coşar, S., Coppola, C., Bellotto, N.: Volume-based human re-identification with RGB-d cameras. In: Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP) - Volume 4: VISAPP, pp. 389–397 (2017)
- Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995). <https://doi.org/10.1023/A:1022627411411>
- Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person Re-Identification by Symmetry-Driven Accumulation of Local Features. In: IEEE CVPR, pp. 2360–2367. <https://doi.org/10.1109/CVPR.2010.5539926> (2010)
- Ghiass, R.S., Arandjelović, O., Bendada, H., Maldague, X.: Infrared Face Recognition: a Literature Review. In: The 2013 International Joint Conference on Neural Networks (IJCNN), pp. 1–10. <https://doi.org/10.1109/IJCNN.2013.6707096> (2013)
- Hermosilla, G., del Solar, J.R., Verschae, R., Correa, M.: A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recogn.* **45**(7), 2445–2459 (2012). <https://doi.org/10.1016/j.patcog.2012.01.001>
- Joachims, T., Nédellec, C., Rouveirol, C. (eds.): Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Springer, Berlin (1998)
- Koide, K., Miura, J.: Identification of a specific person using color, height, and gait features for a person following robot. *Robot. Auton. Syst.* **84**, 76–87 (2016). <https://doi.org/10.1016/j.robot.2016.07.004>
- Kullback, S., Leibler, R.A.: On information and sufficiency. *Ann. Math. Statist.* **22**(1), 79–86 (1951). <https://doi.org/10.1214/aoms/1177729694>
- Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(7), 1622–1634 (2013). <https://doi.org/10.1109/TPAMI.2012.246>
- Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: Deep Filter Pairing Neural Network for Person Re-Identification. In: IEEE CVPR (2014)
- Mensink, T., Verbeek, J., Perronnin, F., Csorika, G.: Distance-based image classification: Generalizing to new classes at near-zero cost. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(11), 2624–2637 (2013). <https://doi.org/10.1109/TPAMI.2013.83>
- Møgelmoose, A., Bahnsen, C., Moeslund, T.B., Clapès, A., Escalera, S.: Tri-Modal Person Re-Identification with Rgb, Depth and Thermal Features. In: IEEE CVPR Workshops 2013, Portland, pp. 301–307 (2013)
- Munaro, M., Basso, A., Fossati, A., Gool, L.V.: Menegatti, E.: 3D Reconstruction of Freely Moving Persons for Re-Identification with a Depth Sensor. In: 2014 IEEE ICRA, pp. 4512–4519. <https://doi.org/10.1109/ICRA.2014.6907518> (2014)
- Munaro, M., Fossati, A., Basso, A., Menegatti, E., Van Gool, L.: One-shot person re-identification with a consumer depth camera. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) *Person Re-Identification*, pp. 161–181. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6296-4_8

22. Nguyen, D.T., Hong, H.G., Kim, K.W., Park, K.R.: Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors* **17**, 605 (2017)
23. Olesen, B.W.: Thermal comfort. *Techn. Rev.* **2**, 3–37 (1982)
24. Paisitkriangkrai, S., Shen, C., van den Hengel, A.: Learning to Rank in Person Re-Identification with Metric Ensembles. In: *IEEE CVPR* (2015)
25. Pala, F., Satta, R., Fumera, G., Roli, F.: Multimodal person reidentification using rgb-d cameras. *IEEE Trans. Circ. Syst. Video Technol.* **26**(4), 788–799 (2016)
26. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv.* **46**(2), 29:1–29:37 (2013)
27. W, B.: The temperature of the skin surface. *J. Am. Med. Assoc.* **106**(14), 1158–1162 (1936). <https://doi.org/10.1001/jama.1936.02770140020007x>
28. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and Appearance Context Modeling. In: *IN: PROC. ICCV* (2007)
29. Weinrich, C., Volkhardt, M., Gross, H.M.: Appearance-Based 3D Upper-Body Pose Estimation and Person Re-Identification on Mobile Robots. In: *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4384–4390. <https://doi.org/10.1109/SMC.2013.748> (2013)
30. Wengelfeld, T., Eisenbach, M., Trinh, T.Q., Gross, H.M.: May i be your personal coach? bringing together person tracking and visual re-identification on a mobile robot. *ISR 2016* (2016)
31. Wu, Z., Peng, M., Chen, T.: Thermal Face Recognition Using Convolutional Neural Network. In: *2016 International Conference on Optoelectronics and Image Processing (ICOIP)*, pp. 6–9. <https://doi.org/10.1109/OPTIP.2016.7528489> (2016)
32. Yan, Z., Duckett, T., Bellotto, N.: Online learning for human classification in 3d lidar-based tracking. In: *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2017)
33. Zhao, Y., Zhao, X., Xiang, Z., Liu, Y.: Online learning of dynamic multi-view gallery for person re-identification. *Multimed. Tools Appl.* **76**(1), 217–241 (2017). <https://doi.org/10.1007/s11042-015-3015-5>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Serhan Coşar is a post-doctoral research fellow at Lincoln Centre for Autonomous Systems Research (L-CAS), University of Lincoln, United Kingdom. He received his M.Sc. and Ph.D. degrees of Electronics Engineering and Computer Science from Sabancı University, Istanbul, Turkey in 2008 and 2013, respectively. He is author of 3 book chapter, 7 international journal papers (CVIU,IMAVIS) and 12 international conference papers (ACPR,ICCV,AVSS) and he has participated in a number of European and National projects. His research interest includes robot perception, human tracking, activity recognition, sparse representation, and distributed estimation.

Nicola Bellotto is a Reader (UK Associate Professor) in the School of Computer Science, University of Lincoln, and a member of the Lincoln Centre for Autonomous Systems (L-CAS). His main research interests are in machine perception, especially for human detection, tracking, identification and activity recognition with autonomous mobile robots. He has a Master in Electronic Engineering from the University of Padua, Italy, and a PhD in Computer Science from the University of Essex, UK. Before joining the University of Lincoln, Dr Bellotto was a researcher in the Active Vision Lab at the University of Oxford. He is a Senior Member of IEEE and an Executive Committee Member of the British Machine Vision Association (BMVA). Dr Bellotto is the recipient of a Google Faculty Research Award and a PI/Co-I in several EU and UK projects on autonomous mobile robots.